



Reduced bias estimation of the log odds ratio

Asma Saleh¹

Received: 21 November 2022 / Revised: 2 July 2024 / Published online: 26 August 2024
© Crown 2024

Abstract

Analysis of binary matched pairs data is problematic due to infinite maximum likelihood estimates of the log odds ratio and potentially biased estimates, especially for small samples. We propose a penalised version of the log-likelihood function based on adjusted responses which always results in a finite estimator of the log odds ratio. The probability limit of the adjusted log-likelihood estimator is derived and it is shown that in certain settings the maximum likelihood, conditional and modified profile log-likelihood estimators drop out as special cases of the former estimator. We implement indirect inference to the adjusted log-likelihood estimator. It is shown, through a complete enumeration study, that the indirect inference estimator is competitive in terms of bias and variance in comparison to the maximum likelihood, conditional, modified profile log-likelihood and Firth's penalised log-likelihood estimators.

Keywords Bias reduction · Binary matched pairs · Indirect inference · Maximum likelihood · Modified profile likelihood · Adjusted responses

Mathematics Subject Classification 62F10 · 62F12

1 Introduction

Consider a series of q independent pairs of independent binomial random variables (Y_{i1}, Y_{i2}) , with $Y_{i1} \sim \text{Bi}(1, \pi_{i1})$ and $Y_{i2} \sim \text{Bi}(m_i, \pi_{i2})$ as in Section 4 of Lunardon (2018). Let the success probabilities satisfy $\pi_{i1} = \exp(\psi + \lambda_i) / \{1 + \exp(\psi + \lambda_i)\}$ and $\pi_{i2} = \exp(\lambda_i) / \{1 + \exp(\lambda_i)\}$, where $\psi = \log\{\pi_{i1} / (1 - \pi_{i2})\} - \log\{\pi_{i2} / (1 - \pi_{i2})\}$, the log odds ratio, is the parameter of interest and $\lambda_i = \log\{\pi_{i2} / (1 - \pi_{i2})\}$ is the nuisance parameter, $i = 1, \dots, q$. This is a stratified setting as in Sartori (2003) where the sample size is $n = \sum_{i=1}^q m_i$ and where q is the number of strata and m_i is the i th stratum sample size. We will refer to this exponential family model in canonical

✉ Asma Saleh
asma.saleh.15@ucl.ac.uk

¹ Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK

form as the binomial matched pairs model, and to the model with $m_i = m = 1$ as the binary matched pairs model. This model often arises in case–control studies in medical contexts, where y_{i1} and y_{i2} may represent for example the numbers of exposed or experimental persons among one case and m_i controls in the i th stratum and where interest lies in studying the influence of some risk factor or the effect of some treatment. For example, suppose that we have data from a clinical trial evaluating the effectiveness of a new Covid vaccine in preventing Covid disease. The data was collected according to a matched case–control design with one case and m_i controls per stratum, where $i = 1, \dots, 10$, is the number of strata and m_i , which is the number of control patients, takes on various values from 3 to 7. On each of 10 days patients testing negative for Covid at a specified time in a hospital were served as subjects. On each day one patient chosen at random formed the experimental group and the remainder were controls. The binary response was whether the patient tested positive or not at the end of a specified period, where testing negative is taken as a “success” and the observed numbers y_{i1} and y_{i2} are therefore the numbers of patients in the two groups testing negative for Covid. The object of the analysis is to assess the effect of the drug on the probabilities of success π_{i1} and π_{i2} in the case and control groups, respectively. The parameter of interest ψ is the difference in the log odds ratio of the case and control groups (for more examples, see Section 1.2 of Cox and Snell 1989).

It is well known, since Neyman and Scott (1948), that in stratified modelling settings the maximum likelihood estimator, derived from the profile log-likelihood, is not in general, a consistent or unbiased estimator of the parameter of interest as the dimension of the nuisance parameter increases while the stratum sample size is kept fixed; this is known as the incidental parameter problem. It is possible to solve this problem, in some cases when the model has a particular structure, like in exponential families in canonical form, using conditional or marginal log-likelihoods (see Sections 4.4 and 4.5 of Pace and Salvan 1997). However, these are not always available and so an alternative is to work with the approximate conditional profile log-likelihood of Cox and Reid (1987), which requires orthogonality of the parameter of interest and the nuisance parameter, or the modified profile log-likelihood of Barndorff-Nielson (1983) whose computation requires a sample space derivative. These are often simple to compute for exponential and composite group families and only involve the observed information matrix for the components of the nuisance parameter which is readily available from direct differentiation (see Section 4.7 of Pace and Salvan 1997) and often provide accurate approximations to conditional or marginal log-likelihoods when they exist. It has been shown through many examples that when the profile log-likelihood performs poorly, approximate conditional and modified profile log-likelihoods can perform much better (see Section 3.1, Example 1 of McCullagh and Tibshirani 1990).

An alternative family of estimators in regular parametric problems was developed in Firth (1993) where the first order bias term is removed from the asymptotic bias of the maximum likelihood estimator by solving a set of adjusted score equations. Firth (1993) considered the case of exponential families with canonical parametrisation, amongst others, and showed that for this family of models, his method is equivalent to maximising a penalised likelihood where the penalty function is the Jeffreys invariant prior. Lunardon (2018) showed that the bias reduction approach of Firth (1993) provides an inferential framework which is, from an asymptotic perspective, equivalent to

that for the approximate conditional and modified profile log-likelihoods when dealing with nuisance parameters. The advantage of bias reduction of Firth (1993) is that it can handle the problem of monotone likelihoods for stratified models with categorical responses. Nevertheless, the approach of Firth (1993) is not in general invariant under interest-respecting reparameterizations.

Indirect inference is another class of inferential procedures that appeared in the Econometrics literature in Gourieroux et al. (1993) and has been used for bias reduction of the maximum likelihood estimator and of other estimators. Its simplest version proceeds by subtracting from the maximum likelihood estimator its full bias and evaluating it at the new indirect inference estimator which therefore becomes the solution of an implicit equation. Kuk (1995) describes a simulation based approach for indirect inference by implementing an iterative bias correction of any suitably defined initial estimator, to yield an estimator which is asymptotically unbiased and consistent.

In this paper, we review in Sect. 2 the profile, conditional, modified profile and Firth (1993) penalised likelihood estimators of the log odds ratio. Since the maximum likelihood, conditional and modified profile likelihood estimators inherit the problem of infinite estimates of the log odds ratio, we propose in Sect. 3 a penalised log-likelihood function based on adjusted responses which always yields finite point estimates of the parameter of interest. The probability limit of the adjusted log-likelihood estimator is derived and it is shown that in certain settings the maximum likelihood, conditional and modified profile log-likelihood estimators drop out as special cases of the former estimator. In Sect. 4 we implement indirect inference to reduce the bias of the adjusted log-likelihood estimator, by adapting the method of Kuk (1995) to nuisance parameter settings. The finite-sample properties of all the above estimators are compared through a complete enumeration study in Sect. 5, as in Lunardon (2018) where no simulation is required, followed by a discussion. Finally, a real data set is analysed in Sect. 6.

2 Review of point estimation of the log odds ratio

2.1 Maximum likelihood

Several estimators of the common log odds ratio ψ have been proposed in the literature. These include the Mantel and Haenszel, empirical logit and Birch estimators (see Breslow 1981; Gart 1971, for a review of these estimators and of their properties). In this section, however, we only consider estimators of the log odds ratio that depend on the data only through the sufficient statistic.

The log-likelihood function for $\theta = (\psi, \lambda_1, \dots, \lambda_q)^\top$ for the above binomial matched pairs model is (Lunardon 2018, Section 4.1)

$$l(\theta) = \sum_{i=1}^q \psi y_{i1} + \sum_{i=1}^q \lambda_i (y_{i1} + y_{i2}) - \sum_{i=1}^q [\log\{1 + \exp(\psi + \lambda_i)\} + m_i \log\{1 + \exp(\lambda_i)\}]. \quad (1)$$

This is a linear full exponential family in canonical form (Pace and Salvan 1997, Section 5) with jointly sufficient statistics $t = \sum_{i=1}^q y_{i1}$ and $s_i = y_{i1} + y_{i2}$ for ψ and λ_i , respectively. Throughout this section, we consider for simplicity $m_i = m$ with totals $s_i = (m + 1)/2$ as in (Sartori 2003, Example 3). In this setting, the constrained maximum likelihood estimator of λ_i for a fixed value of ψ , denoted by $\hat{\lambda}_{i,\psi}$, will be identical for all $i = 1, \dots, q$ and so we set $\hat{\lambda}_{i,\psi} = \hat{\gamma}_\psi$. Equivalently, denote the constrained maximum likelihood estimator of ψ for a fixed value of γ by $\hat{\psi}_\gamma$. The score equations for the log-likelihood function with respect to γ and ψ are respectively

$$\frac{m+1}{2} - \frac{\exp(\psi + \gamma)}{1 + \exp(\psi + \gamma)} - m \frac{\exp(\gamma)}{1 + \exp(\gamma)} = 0, \quad (2)$$

$$t - q \frac{\exp(\psi + \gamma)}{1 + \exp(\psi + \gamma)} = 0. \quad (3)$$

The solution of (3) is $\hat{\psi}_\gamma = \log\{t/(q - t)\} - \gamma$, and on substituting this in (2) and solving for γ we get the maximum likelihood estimator of the nuisance parameter $\hat{\gamma} = \log[\{q(m+1) - 2t\}/\{q(m-1) + 2t\}]$. Substituting $\hat{\gamma}$ in $\hat{\psi}_\gamma$ we get the maximum likelihood estimator of the parameter of interest

$$\hat{\psi} = \log\left(\frac{t\{q(m-1) + 2t\}}{\{q-t\}\{q(m+1) - 2t\}}\right). \quad (4)$$

Note that when $t = 0$ or $t = q$, $\hat{\psi}$ is $-\infty$ or $+\infty$, respectively. This is problematic because it means in these extreme situations were all cases ‘succeed’ or ‘fail’ we cannot estimate ψ . Since ψ is defined as a logarithm of odds ratio, infinite estimates arise when the argument of the logarithm, i.e. the odds ratio is zero. Therefore, a better estimator of ψ that avoids infinity needs to avoid zero as a possibility for the argument of the log function.

Using the weak law of large numbers, Slutsky’s theorem and the Continuous mapping theorem (Florescu 2014, Section 7), we find that $\hat{\psi}$ converges in probability to $\psi + \log[\{(m+1)\exp(\psi) + m - 1\}/\{(m-1)\exp(\psi) + m + 1\}]$ as $q \rightarrow \infty$, and so it is inconsistent. When m is also allowed to increase to ∞ , $\hat{\psi}$ will tend to ψ . This means that $\hat{\psi}$ will be consistent only when both m and q diverge.

Given that the totals s_i are fixed, the maximum likelihood estimator of ψ depends on the data only through the sufficient statistic $T = \sum_{i=1}^q Y_{i1}$ and so its bias and variance can be calculated exactly using

$$E_\psi\{\hat{\psi}(T)\} = \sum_{t=1}^{q-1} \hat{\psi}(t) \frac{\text{pr}(T = t | S_i = s_i)}{1 - \text{pr}(T = 0 | S_i = s_i) - \text{pr}(T = q | S_i = s_i)}, \quad (5)$$

$$\text{var}_\psi\{\hat{\psi}(T)\} = E_\psi[\{\hat{\psi}(T)\}^2] - [E_\psi\{\hat{\psi}(T)\}]^2. \quad (6)$$

2.2 Conditional maximum likelihood

The conditional log-likelihood function is based on the distribution of Y_{i1} given $S_i = s_i$ in each stratum. Davison (1988) and Gart (1970) noted that the conditional density of Y_{i1} given S_i is

$$\text{pr}(Y_{i1} = y_{i1} | S_i = s_i) = \frac{\binom{1}{y_{i1}} \binom{m_i}{s_i - y_{i1}} \exp(\psi y_{i1})}{\sum_{u=0}^{\min(1, s_i)} \binom{1}{u} \binom{m_i}{s_i - u} \exp(\psi u)}. \quad (7)$$

This is the noncentral hypergeometric distribution which is obtained by rewriting the left hand side of (7) as $\text{Pr}(Y_{i1} = y_{i1}, S_i = s_i) / \text{Pr}(S_i = s_i) = \text{Pr}(Y_{i1} = y_{i1}) \text{Pr}(Y_{i2} = s_i - y_{i1}) / \text{Pr}(Y_{i1} + Y_{i2} = s_i)$, by independence of Y_{i1} and Y_{i2} , and noting that $\text{Pr}(Y_{i1} + Y_{i2} = s_i) = \sum_{u=0}^{s_i} \text{Pr}(Y_{i1} = u) \text{Pr}(Y_{i2} = s_i - u)$. Because y_{i1} can only be 0 or 1, the right hand side of (7) can be further simplified to

$$\text{pr}(Y_{i1} = y_{i1} | S_i = s_i) = \left[\frac{s_i \exp(\psi)}{m_i + 1 + s_i \{\exp(\psi) - 1\}} \right]^{y_{i1}} \left[\frac{m_i - s_i + 1}{m_i + 1 + s_i \{\exp(\psi) - 1\}} \right]^{1 - y_{i1}}. \quad (8)$$

This shows that $Y_{i1} | S_i$ has a Bernoulli distribution with success probability the first term inside the bracket of the right hand side of (8). Taking the logarithm of the product of (8) gives the conditional log-likelihood function which simplifies to

$$l_c(\psi) = \sum_{i=1}^q \psi y_{i1} - \sum_{i=1}^q \log [m_i + 1 + s_i \{\exp(\psi) - 1\}]. \quad (9)$$

Letting $m_i = m$, $s_i = (m+1)/2$ and differentiating (9) gives the conditional maximum likelihood estimator

$$\hat{\psi}_c = \log \left(\frac{t}{q - t} \right). \quad (10)$$

When $t = 0$ or $t = q$, $\hat{\psi}_c$ is $-\infty$ or $+\infty$, respectively. In the setting of Lunardon (2018), i.e. when $m_i = m$ and $s_i = (m+1)/2$, the success probability of the Bernoulli random variable $Y_{i1} | S_i$ simplifies to $\pi = \exp(\psi) / \{\exp(\psi) + 1\}$. The distribution of the sufficient statistic T given S_i is therefore Binomial with denominator q and success probability π . The conditional distribution of T can also be obtained using the convolution method following (Butler and Stephens 2017, Section 2). This will be particularly useful for general m_i and s_i where the Binomial conditional distribution of T no longer holds. In fact, the conditional distribution of T will be Poisson binomial.

By noting that T converges in probability to $q\pi$ by the weak law of large numbers, we find that $\log\{t/(q-t)\} \xrightarrow{P} \psi$ by Slutsky's theorem and the Continuous mapping theorem, so $\hat{\psi}_c$ is consistent. As the conditional maximum likelihood estimator depends on the data only through the sufficient statistic, its bias and variance can be calculated using (5) and (6) but replacing $\hat{\psi}(T)$ by $\hat{\psi}_c(T)$.

2.3 Modified profile maximum likelihood

(Davison 2003, Section 12) showed that for a linear exponential family in canonical form, the modified profile log-likelihood function of Barndorff-Nielsen (1983) reduces to

$$l_{mp}(\psi) = l(\psi, \hat{\lambda}_\psi) + \frac{1}{2} \log \left\{ \det j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi) \right\}, \quad (11)$$

where $l(\psi, \hat{\lambda}_\psi)$ is the profile log-likelihood obtained from (1) by substituting the constrained maximum likelihood estimator of λ and where $j_{\lambda\lambda}(\psi, \lambda)$ is the observed information per observation for the λ components and is given by the negative of the second derivative of the log-likelihood function with respect to λ . In the setting $m_i = m$ and $s_i = (m + 1)/2$, $j_{\lambda\lambda}(\psi, \lambda)$ becomes the $q \times q$ matrix with i th diagonal element

$$-\frac{\partial^2 l(\psi, \gamma)}{\partial \gamma^2} = \frac{\exp(\psi + \gamma)}{(1 + \exp(\psi + \gamma))^2} + m \frac{\exp(\gamma)}{(1 + \exp(\gamma))^2}, \quad (12)$$

and zero elsewhere, and where we observed in Sect. 2.1 that the solution of (2), $\hat{\lambda}_\psi = \hat{\gamma}_\psi$, is not available in closed form. This means that there is no closed form expression for (11) and we calculate the maximum modified profile log-likelihood estimator, $\hat{\psi}_{mp}$, numerically and evaluate its bias and variance using (5) and (6), respectively, by replacing $\hat{\psi}(T)$ with $\hat{\psi}_{mp}(T)$.

2.4 Firth penalized likelihood

When $\theta = (\psi, \lambda_1, \dots, \lambda_q)$ is the canonical parameter of an exponential family model like in the model considered here, Firth (1993) showed that the adjusted score equations estimator of θ is equivalent to the maximiser of the penalised log-likelihood function

$$l_*(\theta) = l(\theta) + \frac{1}{2} \log \{\det i(\theta)\}, \quad (13)$$

where $i(\theta) = E\{j(\theta)\}$ is the Fisher information matrix. In the setting $m_i = m$ and $s_i = (m + 1)/2$, the second order partial derivatives of $l(\psi, \lambda_i)$ are

$$\frac{\partial^2 l(\psi, \lambda_i)}{\partial \psi^2} = - \sum_{i=1}^q \frac{\exp(\psi + \lambda_i)}{(1 + \exp(\psi + \lambda_i))^2} \quad (14)$$

$$\frac{\partial^2 l(\psi, \lambda_i)}{\partial \psi \partial \lambda_i} = - \frac{\exp(\psi + \lambda_i)}{(1 + \exp(\psi + \lambda_i))^2} \quad (15)$$

$$\frac{\partial^2 l(\psi, \lambda_i)}{\partial \lambda_i^2} = - \frac{\exp(\psi + \lambda_i)}{(1 + \exp(\psi + \lambda_i))^2} - m \frac{\exp(\lambda_i)}{(1 + \exp(\lambda_i))^2}. \quad (16)$$

Since the above derivatives do not depend on the data, the Fisher information matrix coincides with the observed information and is given by

$$i(\psi, \lambda_i) = \begin{pmatrix} \sum_{i=1}^q V_{i1} & V_{11} & \cdots & V_{q1} \\ V_{11} & (V_{11} + V_{12}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ V_{q1} & 0 & \cdots & (V_{q1} + V_{q2}) \end{pmatrix}, \quad (17)$$

where $V_{i1} = (\exp(\psi + \lambda_i))/(1 + \exp(\psi + \lambda_i))^2$ and $V_{i2} = m(\exp(\lambda_i))/(1 + \exp(\lambda_i))^2$, $i = 1, \dots, q$. The determinant of $i(\psi, \lambda_i)$ is obtained using the standard identity (see, Magnus and Neudecker 2019, Chapter 1, page 28) and simplifies to

$$\det i(\psi, \lambda_i) = \left(\prod_{i=1}^q (V_{i1} + V_{i2}) \right) \left(\sum_{i=1}^q \frac{V_{i1} V_{i2}}{V_{i1} + V_{i2}} \right). \quad (18)$$

Therefore the penalty function that needs to be added to the log-likelihood function is

$$\frac{1}{2} \log\{\det i(\psi, \lambda_i)\} = \frac{1}{2} \sum_{i=1}^q \log(V_{i1} + V_{i2}) + \frac{1}{2} \log \left(\sum_{i=1}^q \frac{V_{i1} V_{i2}}{V_{i1} + V_{i2}} \right). \quad (19)$$

The score equations for the penalised log-likelihood of Firth (1993), $l_*(\psi, \lambda_i)$, with respect to λ_i and ψ involve cumbersome expressions and have no closed form solution so the penalised log-likelihood of Firth (1993) estimator of ψ , denoted by $\hat{\psi}_*$ is obtained numerically. This estimator is always finite as shown in Section 2.1 of Kosmidis and Firth (2021). The bias and variance of $\hat{\psi}_*$ are calculated using (5) and (6), respectively.

2.5 Binary matched pairs model

The binary matched pairs model is a special case of the binomial matched pairs model when $m = 1$. This implies that in the setting of Lunardon (2018), $s_i = 1$, and so $a = d = 0$, where a, b, c and d denote the number of pairs of the form $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$ respectively, with $a + b + c + d = q$. Note that $\sum_{i=1}^q y_{i1} = c + d$, $\sum_{i=1}^q y_{i2} = b + d$ and $\sum_{i=1}^q (y_{i1} + y_{i2}) = b + c + 2d$. We call pairs of the form $(0, 0)$ and $(1, 1)$ concordant, while pairs of the form $(0, 1)$ and $(1, 0)$ are called discordant. In this case, $\hat{\gamma}_\psi = -\psi/2$ and so the profile log-likelihood for ψ is (see Davison 2003, Example 12.23)

$$l_p(\psi) = \psi t - 2q \log\{1 + \exp(\psi/2)\}, \quad (20)$$

which is maximised at

$$\hat{\psi} = 2 \log \left(\frac{t}{q - t} \right)$$

$$= 2 \log \left(\frac{c}{b} \right). \quad (21)$$

Alternatively, $\hat{\psi}$ can be obtained by substituting $m = 1$ in (4). Davison (2003) showed that $\hat{\psi}$ converges in probability to 2ψ as $q \rightarrow \infty$, thus it is inconsistent.

Only discordant pairs enter the conditional log-likelihood and it is given by

$$l_c(\psi) = c\psi - (b + c) \log\{\exp(\psi) + 1\}, \quad (22)$$

which is maximised at

$$\hat{\psi}_c = \log \left(\frac{c}{b} \right), \quad (23)$$

which converges in probability to ψ as $q \rightarrow \infty$, as noted by (Davison 2003, Example 12.23), so it is consistent.

Substituting $\hat{\gamma}_\psi = -\psi/2$ in (12) gives $j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi) = 2(\exp(\psi/2))/(1 + \exp(\psi/2))^2$, and in this case (Davison 2003, Example 12.23) showed that

$$l_{mp}(\psi) = \frac{1}{4}\psi(q + 4t) - 3q \log\{1 + \exp(\psi/2)\}, \quad (24)$$

which is maximized at

$$\begin{aligned} \hat{\psi}_{mp} &= 2 \log \left(\frac{q + 4t}{5q - 4t} \right) \\ &= 2 \log \left(\frac{b + 5c}{c + 5b} \right), \end{aligned} \quad (25)$$

where the latter converges in probability to $2 \log \left[\frac{1 + 5 \exp(\psi)}{5 + \exp(\psi)} \right]$ as $q \rightarrow \infty$. Note that when $c = 0$ or $b = 0$, $\hat{\psi}_{mp}$ is $2 \log(1/5)$ or $2 \log(5)$, respectively, i.e. $\hat{\psi}_{mp}$ is finite. Although $\hat{\psi}_{mp}$ is inconsistent, Davison (2003) showed that it is less biased than $\hat{\psi}$.

3 Adjusted likelihood method

3.1 Penalised likelihood based on adjusted responses

In order to avoid infinite estimates of ψ , as is the case with $\hat{\psi}$, $\hat{\psi}_c$ and $\hat{\psi}_{mp}$ (for $m \neq 1$), when all of the y_{i1} observations are zero or one, we propose to adjust the log-likelihood function by adding a small number $\delta > 0$ to each success, y_{i1} and y_{i2} , and to each failure, $1 - y_{i1}$ and $m_i - y_{i2}$. The penalised log-likelihood function based on adjusted responses for $\theta = (\psi, \lambda_1, \dots, \lambda_q)^\top$ becomes

$$l_a(\theta) = \sum_{i=1}^q \psi(y_{i1} + \delta) + \sum_{i=1}^q \lambda_i(y_{i1} + y_{i2} + 2\delta)$$

$$- \sum_{i=1}^q [(1 + 2\delta) \log\{1 + \exp(\psi + \lambda_i)\} + (m_i + 2\delta) \log\{1 + \exp(\lambda_i)\}]. \quad (26)$$

When $m_i = m$ and $s_i = (m + 1)/2$, the score equations for the above log-likelihood function with respect to γ and ψ simplify respectively to

$$\frac{m + 1 + 4\delta}{2} - (1 + 2\delta) \frac{\exp(\psi + \gamma)}{1 + \exp(\psi + \gamma)} - (m + 2\delta) \frac{\exp(\gamma)}{1 + \exp(\gamma)} = 0, \quad (27)$$

$$t + q\delta - q(1 + 2\delta) \frac{\exp(\psi + \gamma)}{1 + \exp(\psi + \gamma)} = 0, \quad (28)$$

where we set the constrained penalised maximum likelihood estimator of λ_i , based on adjusted responses, for a fixed value of ψ , denoted by $\hat{\lambda}_{i,\psi,a}$, to be $\hat{\gamma}_{\psi,a}$ because it will be identical for all $i = 1, \dots, q$. The simultaneous solution of (27) and (28) give the penalised maximum likelihood estimators of γ and ψ , based on adjusted responses

$$\hat{\gamma}_a = \log \left[\frac{q(m + 1 + 2\delta) - 2t}{q(m - 1 + 2\delta) + 2t} \right], \quad (29)$$

$$\hat{\psi}_a = \log \left[\frac{(t + q\delta)\{q(m - 1 + 2\delta) + 2t\}}{\{q(m + 1 + 2\delta) - 2t\}\{q(1 + \delta) - t\}} \right]. \quad (30)$$

Note that when $\delta = 0$, $\hat{\psi}_a = \hat{\psi}$. Note also that when $t = 0$ or $t = q$, $\hat{\psi}_a$ is finite, while when $t = q/2$, $\hat{\psi}_a = 0$. When $m = 1$,

$$\hat{\psi}_a = 2 \log \left[\frac{t + q\delta}{q(1 + \delta) - t} \right]. \quad (31)$$

Since $\hat{\psi}_a$ depends on the data only through the sufficient statistic t , its bias and variance are computed using (5) and (6), respectively.

3.2 Probability limit of the penalised likelihood estimator based on adjusted responses

In this section we obtain the probability limit of the penalised log-likelihood estimator based on adjusted responses and derive the relationship that δ should satisfy in order to make this estimator consistent. We also show how the modified profile log-likelihood estimator (when $m = 1$) and the conditional log-likelihood estimator can be recovered for particular values of δ .

When $m = 1$, as $q \rightarrow \infty$, $\hat{\psi}_a$ converges in probability to

$$2 \log \left(\frac{\delta \{\exp(\psi) + 1\} + \exp(\psi)}{\delta \{\exp(\psi) + 1\} + 1} \right), \quad (32)$$

while for a general m , we find that as $q \rightarrow \infty$, $\hat{\psi}_a$ converges in probability to

$$\log \left(\frac{[\delta \{\exp(\psi) + 1\} + \exp(\psi)][\{m - 1 + 2\delta\} \{\exp(\psi) + 1\} + 2 \exp(\psi)]}{[\delta \{\exp(\psi) + 1\} + 1][\{m + 1 + 2\delta\} \{\exp(\psi) + 1\} - 2 \exp(\psi)]} \right). \quad (33)$$

Similar to Table 12.3 of Davison (2003), Table 1 compares the limiting values of $\hat{\psi}$, $\hat{\psi}_c$, $\hat{\psi}_{mp}$ and $\hat{\psi}_a$ when $m = 1$ for a set of values of ψ ranging from 0 to 5 and a set of values of δ ranging from 0.05 to 0.50. We note that for any given ψ , there exists a value of δ for which the limit of $\hat{\psi}_a$ is closer to the truth than $\hat{\psi}_{mp}$. In other words, there is evidence that there exists a δ value such that $\hat{\psi}_a$ converges to the truth faster than $\hat{\psi}_{mp}$. These values of δ decrease as the true value of ψ increase. Observe also that when $\delta = 0.25$, the limiting value of $\hat{\psi}_a$ coincides with that of $\hat{\psi}_{mp}$. In fact substituting $\delta = 0.25$ in (32), we find that $\hat{\psi}_a$ converges in probability to the same limit of $\hat{\psi}_{mp}$ given in Sect. 2.5. This means that the penalised log-likelihood estimator based on adjusted responses recovers the modified profile log-likelihood estimator when $m = 1$ and $\delta = 0.25$, i.e. $\hat{\psi}_a = \hat{\psi}_{mp}$.

When $m = 1$, in order to make $\hat{\psi}_a$ consistent we need to equate the ratio inside the logarithm of (32) with $\sqrt{\exp(\psi)}$ which simplifies to the equation

$$[\exp(\psi) - 1][\delta^2 \{\exp(\psi)\}^2 + \{2\delta^2 - 1\} \{\exp(\psi)\} + \delta^2] = 0. \quad (34)$$

When $\psi = 0$, there is no adjustment because there is no bias so we consider the positive solution of the quadratic equation $[\delta^2 \{\exp(\psi)\}^2 + \{2\delta^2 - 1\} \{\exp(\psi)\} + \delta^2] = 0$ in terms of δ which simplifies to

$$\delta = \frac{\sqrt{\exp(\psi)}}{1 + \exp(\psi)}. \quad (35)$$

Substituting (35) into (31) and solving for ψ gives the same estimate as $\hat{\psi}_c$. This means that the value of δ that achieves consistency of $\hat{\psi}_a$ is the one that recovers $\hat{\psi}_c$. This is disadvantageous because we inherit exactly the same problems with conditional log-likelihood (i.e. infinite estimates) if we attempt to tune δ to make $\hat{\psi}_a$ consistent. The value of δ in terms of t and q that recovers $\hat{\psi}_c$ is obtained by equating (31) with $\hat{\psi}_c$ and simplifies to

$$\delta = \sqrt{\frac{t(3qt - 2t^2 - q^2)}{q^2(2t - q)}}. \quad (36)$$

Observe that when $t = 0$ or $t = q$, $\delta = 0$ and so $\hat{\psi}_a = \hat{\psi}$, while when $t = q/2$, δ is infinite.

For a general m , the relationship that δ should satisfy in order to make $\hat{\psi}_a$ consistent is found by equating the ratio inside the logarithm of (33) with $\exp(\psi)$ which simplifies to the equation

$$\delta \{ \exp(\psi) \}^3 \{ m - 1 + 2\delta \} - \{ \exp(\psi) \}^2 \{ 2 + \delta(1 - m - 2\delta) \}$$

Table 1 Probability limits of profile, conditional, modified profile and adjusted log-likelihood based on adjusted responses estimators of the log odds ratio ψ in the binary matched pairs model when $m = 1$. For each ψ , the value in bold face corresponds to the limiting value closest to the truth if we ignore $\hat{\psi}_c$

ψ		0	0.5	1	1.5	2.0	2.5	3	4	5
Limit of $\hat{\psi}$		0	1	2	3	4	5	6	8	10
Limit of $\hat{\psi}_c$		0	0.5	1	1.5	2	2.5	3	4	5
Limit of $\hat{\psi}_{mp}$		0	0.66	1.27	1.81	2.24	2.56	2.79	3.05	3.16
Limit of $\hat{\psi}_a$	δ	0.05	0	0.91	1.79	2.63	3.41	4.09	4.66	5.83
		0.10	0	0.83	1.62	2.36	3.00	3.52	3.93	4.65
		0.15	0	0.76	1.49	2.14	2.69	3.12	3.44	3.82
		0.20	0	0.71	1.37	1.96	2.44	2.81	3.08	3.51
		0.25	0	0.66	1.27	1.81	2.24	2.56	2.79	3.05
		0.30	0	0.62	1.19	1.68	2.07	2.36	2.56	2.79
		0.35	0	0.58	1.12	1.57	1.93	2.19	2.37	2.57
		0.40	0	0.55	1.05	1.47	1.81	2.05	2.21	2.39
		0.45	0	0.52	0.99	1.39	1.70	1.92	2.07	2.24
		0.50	0	0.49	0.94	1.32	1.60	1.81	1.95	2.10

$$+\{\exp(\psi)\}\{2 + \delta(1 - m - 2\delta)\} - \delta(m - 1 + 2\delta) = 0, \quad (37)$$

with no closed form solution. The value of δ in terms of t , q and m that recovers $\hat{\psi}_c$ satisfies the equation

$$2q^2\delta^2(q - 2t) + q^2\delta\{q(m - 1) + 2t(1 - m)\} - 2t(q^2 - 3tq + 2t^2) = 0. \quad (38)$$

Fig. 1 shows a plot of δ , the root of (37), against ψ for $m = 1$, $m = 3$, $m = 11$ and $m = 39$. This plot shows that there is a scaled logistic relationship between δ and ψ and that the best choice of δ lies in the range $0 < \delta < 0.5$. Given the true value of ψ , as m increases, the value of δ that makes $\hat{\psi}_a$ consistent decreases to zero. This is expected because we know from standard asymptotic theory that the maximum likelihood estimator is asymptotically consistent.

4 Indirect inference estimation of the log odds ratio

Suppose that $\hat{\psi}$ is some initial estimator of ψ , not necessarily the maximum likelihood estimator, then the simplest method of bias reduction of $\hat{\psi}$ via indirect inference relies on solving the equation (see, Kuk 1995)

$$\tilde{\psi} = \hat{\psi} - B_{\hat{\psi}}(\tilde{\psi}, \lambda), \quad (39)$$

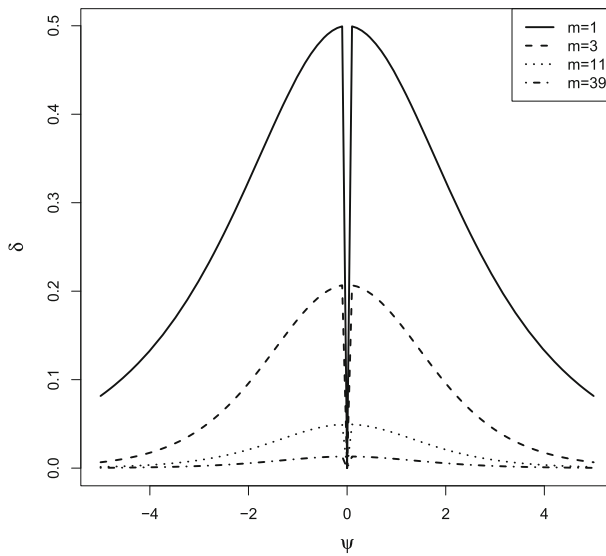


Fig. 1 Plot of the value of δ , that makes the penalised log-likelihood based on adjusted responses estimator, $\hat{\psi}_a$, consistent, against ψ for $m = 1, m = 3, m = 11$ and $m = 39$

with respect to $\tilde{\psi}$ where $B_{\tilde{\psi}}(\tilde{\psi}, \lambda) = E_{\tilde{\psi}, \lambda}(\hat{\psi}) - \tilde{\psi}$ is the bias function of $\hat{\psi}$ evaluated at $\tilde{\psi}$ and $\lambda = \lambda_1, \dots, \lambda_q$. We call $\tilde{\psi}$ the indirect inference estimator of ψ . Alternatively, (39) can be written as

$$\hat{\psi} = E_{\tilde{\psi}, \lambda}(\hat{\psi}). \quad (40)$$

Since we want to reduce the bias of $\hat{\psi}_a$ when $m_i = m$ and $s_i = (m + 1)/2$, our indirect inference estimator $\tilde{\psi}$ is the solution of

$$\hat{\psi}_a = E_{\tilde{\psi}, \gamma}(\hat{\psi}_a). \quad (41)$$

As the expectation of $\hat{\psi}_a$ can be obtained using complete enumeration, one version of $\tilde{\psi}$, $\tilde{\psi}_{a*}$, which is independent of γ , can be defined by using the conditional density $\Pr(T = u | S_i; \psi)$ which is binomial with denominator q and success probability $\exp(\psi)/\{\exp(\psi) + 1\}$,

$$\hat{\psi}_a(T) = \sum_{u=0}^q \hat{\psi}_a(u) \Pr(T = u | S_i; \tilde{\psi}_{a*}). \quad (42)$$

No closed form solution exists for $\tilde{\psi}_{a*}$, so we solve the above equation numerically and calculate the expectation and variance of $\tilde{\psi}_{a*}$ using (5) and (6). For $t \in \{0, q\}$, $\tilde{\psi}_{a*}$ has no solution. In fact the expectation in the right hand side of (42) is bounded below by $\hat{\psi}_a(0)$ and above by $\hat{\psi}_a(q)$. This means that when the binary observations are all zero or one the indirect inference estimator is not defined which is unfortunate because even though we overcome the problem of infinite estimates at $t = 0$ and

$t = q$ by introducing a penalised likelihood estimator based on adjusted responses, when we attempt to reduce the bias of the later the same problem appears again at those boundary values of t .

5 Complete enumeration study

We reproduce the complete enumeration study in (Lunardon 2018, Table 1) which compares the finite sample bias and variance of estimators derived from profile, conditional, modified profile and penalized (Firth (1993)) likelihoods, denoted by $\hat{\psi}$, $\hat{\psi}_c$, $\hat{\psi}_{mp}$ and $\hat{\psi}_*$, respectively. We enrich this study by adding the penalised maximum likelihood estimator based on adjusted responses, $\hat{\psi}_a$, and the indirect inference estimator based on $\hat{\psi}_a$ using the conditional model, denoted by $\hat{\psi}_{a*}$, for a set of 20 values of δ ranging from 0.05 to 1.00. The comparison in (Lunardon 2018, Table 1) also assesses the coverage probability and length of 95% confidence intervals for ψ based on the chi-squared approximation to the distribution of $W_*(\psi)$ and to the distributions of the profile, conditional and modified profile log-likelihood ratios, denoted by $W(\psi)$, $W_c(\psi)$ and $W_{mp}(\psi)$, respectively. We extend this comparison by adding the coverage probability and length of 95% confidence intervals for ψ based on the chi-squared approximation to the distribution of the penalised log-likelihood ratio based on adjusted responses, denoted by $W_a(\psi)$. The exact bias and variance of estimators and the exact coverage probability and average length of confidence intervals is obtained through complete enumeration because the estimators and confidence intervals all depend on the sufficient statistic $T = \sum_{i=1}^q Y_{i1}$ and so the distribution of T given $S_1 = s_1, \dots, S_q = s_q$ can be computed numerically following Butler and Stephens (2017). These summaries however, are computed only for $t \in \{1, \dots, q-1\}$ since when $t \in \{0, q\}$, $\hat{\psi}$, $\hat{\psi}_c$ and $\hat{\psi}_{mp}$ (for $m \neq 1$) are infinite.

Tables 5, 6, 7, 8, 9 and 10 report the bias and variance of estimators, while Tables 11, 12, 13, 14, 15 and 16 report the coverage probability and average length of confidence intervals with nominal level 95% when the true log odds ratio ψ is unity, with $m \in \{1, 3, 11, 39\}$ and $q \in \{30, 100, 1000\}$.

Overall, for fixed δ the numerical value of the bias and variance of the estimator $\hat{\psi}_a$ decrease as m increases. In many cases, however this means that the bias becomes more negative, i.e. the magnitude of the bias increases with m . This suggests that for any combination of q and m , there exists a particular value of δ , above which the bias of $\hat{\psi}_a$ does not improve. In fact for any combination of m and q , there exists a value of δ such that $\hat{\psi}_a$ has minimum bias which is smaller than the bias of the estimators $\hat{\psi}$, $\hat{\psi}_c$, $\hat{\psi}_{mp}$ and $\hat{\psi}_*$; for example, for $q = 30$, $m = 1$ this optimal δ in terms of bias is 0.45, for the combinations $q = 30$, $m = 39$ and $q \geq 100$, $m \geq 11$, the optimal value of δ that gives minimum bias becomes smaller than 0.05. For $q = 30$, at the optimum δ value, the estimator $\hat{\psi}_a$ has smaller bias and variance than $\hat{\psi}_c$. This is also true for other values of q , except that it is not very clear from Tables 5, 6, 7, 8, 9 and 10 because we consider a specific set of values for δ ; for example, for $q = 100$ and $m = 11$, at $\delta = 0.045$, $\hat{\psi}_a$ has bias and variance -0.03 and 0.47 , respectively (both multiplied by 10), while for $q = 1000$ and $m = 1$, at $\delta = 0.444$, $\hat{\psi}_a$ has bias and

variance 0.00 and 0.04, respectively (both multiplied by 10). This optimal value of δ decrease as m increases, but for a fixed m , as q increase above 30, the optimum δ value remain constant. This is expected because the theoretical derivations in equations (35) and (37) tell us that the optimum value of δ in terms of consistency is independent of q . For example, the optimum value of δ when $m = 1$ and $\psi = 1$ is given by (35) and is equal to $\delta = 0.4434$. Similarly, (37) gives the optimum value of δ in terms of consistency for a general value of m and ψ , for example, when $m = 3$ and $\psi = 1$, the optimum δ is 0.1566.

So in terms of the incidental parameter problem, for a fixed value of m and increasing q , there is an optimum value of δ that makes $\hat{\psi}_a$ less biased and has smaller variance than all the other estimators. This behaviour can be seen more clearly from Table 17 where the optimum value of δ that minimizes the bias of $\hat{\psi}_a$ was chosen from a finer set of δ values. For the values of m and q chosen in Table 17, the effect of fixing m and allowing q to increase is a larger optimal δ , while fixing q and increasing m decreases this optimal δ value. However, the pattern in Table 17 suggest that as both q and m are allowed to diverge, the optimal δ value becomes close to zero and this makes intuitive sense because we know that $\delta = 0$ gives rise to the maximum likelihood estimator $\hat{\psi}$ which is asymptotically unbiased when both m and q tend to ∞ .

Nevertheless, the bias results for the estimator $\hat{\psi}_{a*}$ show a marked improvement over $\hat{\psi}$, $\hat{\psi}_{mp}$ and $\hat{\psi}_c$, for all values of q and m . When $\hat{\psi}_{a*}$ is compared with $\hat{\psi}_*$, the former has smaller bias and variance for $m = 1$ and $m = 3$ for all values of q . When compared with $\hat{\psi}_a$, the bias of $\hat{\psi}_{a*}$ is reduced for all values of δ except the optimal one. The interesting result to note here is that as δ increases, the bias and variance of $\hat{\psi}_{a*}$ approach that of $\hat{\psi}_c$ for all combinations of q and m considered. To wrap up this comparison of estimators, if we were to choose a δ based on $\hat{\psi}_{a*}$, then it will be the one that is close to (but not equal to) zero because it will give the smallest bias and variance. However, $\hat{\psi}_{a*}$ does not perform better than $\hat{\psi}_a$ all the time. In fact, if we were to choose a δ based on $\hat{\psi}_a$, then it will be the optimal δ that minimizes the bias of $\hat{\psi}_a$ because the bias and variance of $\hat{\psi}_a$ at this optimal δ is smaller than that of any other estimator in the table for any combination of q and m . Having said that, in practice we are only given a data set and we don't know the particular δ for that data set so $\hat{\psi}_{a*}$ will be a good choice because its bias and variance are very competitive regardless of the value of δ .

Concerning the coverage probability and average length of 95% confidence intervals, Lunardon (2018) noted that intervals derived from $W_{mp}(\psi)$ and $W_*(\psi)$ are consistent with those from $W_c(\psi)$ for $m \geq 3$. The coverage probability and average length of confidence intervals derived from $W_a(\psi)$ show an improvement over those derived from $W_c(\psi)$ for particular values of δ (shown in bold face in Tables 11, 12, 13, 14, 15, 16). We considered 20 values of δ ranging from 0.31 to 0.50 for $m = 1$, 0.06 to 0.25 for $m = 3$ and 0.01 to 0.20 for $m = 11$ and $m = 39$. For those particular values of δ in bold face, the coverage probability derived from $W_a(\psi)$ is closer to the nominal coverage of 95% than that derived from $W_c(\psi)$. This agrees with the fact that $\hat{\psi}_a$ performs better than $\hat{\psi}_c$ in terms of bias and variance for some optimal value of δ . As the value of δ increase the average length of confidence intervals derived from $W_a(\psi)$ decreases which is expected because the variance (and hence standard error) of $\hat{\psi}_a$ becomes smaller as δ becomes larger.

In conclusion it has been shown how, in the binomial matched pairs model, finite estimates of the log odds ratio are produced in cases where the observations are either all equal to zero or all equal to one by penalising the log-likelihood function through the additive adjustment of a tuning parameter $\delta > 0$ to each success and failure. This value δ was then used as a parameter that could be tuned to improve the bias and/or variance of the penalised log-likelihood estimator based on adjusted responses. The indirect inference method was applied to further reduce the bias of $\hat{\psi}_a$. The method can be applied in principle to any parametric model. We are also free to use estimators other than $\hat{\psi}_a$ as initial estimates. Indeed the setting considered here where m_i and s_i are fixed to m and $(m + 1)/2$, respectively is a special case and perhaps a large scale simulation study would be useful to account for different stratum sizes and totals.

6 Analysis of crying babies data

In this section we illustrate the methods discussed above in a general setting by providing a real-data example with different stratum sizes. We re-analyse the crying of babies data set given in (Cox and Snell 1989, Example 1.2). The data come from an experiment intended to assess the effectiveness of rocking motion on the crying of babies and were collected according to a matched case–control design with one case and m_i controls per stratum, where $i = 1, \dots, 18$ and m_i takes on various values from 5 to 9. On each of 18 days babies not crying at a specified time in a hospital were served as subjects. On each day one baby chosen at random formed the experimental group and the remainder were controls. The binary response was whether the baby was crying or not at the end of a specified period. In (Cox and Snell 1989, Example 1.2), not crying is taken as a "success" and the observed numbers y_{i2} and y_{i1} are therefore the numbers of babies in the two groups not crying. The number of non crying babies in the experimental group is $t = 15$.

The estimates of the log odds ratio ψ and its standard error are reported in Table 2. Davison (1988) obtained the maximum likelihood and conditional maximum likelihood estimates while Lunardon (2018) obtained the estimates of ψ derived from the modified profile and penalised (Firth) log-likelihoods. We found the penalised based on adjusted responses log-likelihood and indirect inference estimates for 20 values of δ ranging from 0.05 to 1.00. As the true log odds ratio ψ is unknown, it is difficult to decide which estimator should be preferred however, there is an important observation to note. As δ approaches 1, the indirect inference estimates of ψ approach the conditional log-likelihood estimates and the standard errors of $\hat{\psi}_{a*}$ approach that of $\hat{\psi}_c$. This observation has been noted before in the complete enumeration study in terms of bias and variance in the specific setting where the i th stratum sample size was fixed. Even though this observation has not been proved analytically, the results based on the crying babies data show that, at least numerically, it may also hold for a general binary matched pairs with different stratum sizes.

The standard errors of the estimates in Table 2, except for $\hat{\psi}_c$, are obtained using the Fisher information matrix evaluated at the given estimate of ψ . In particular, for the maximum likelihood estimator the standard error is obtained using

$\sqrt{\text{diag}\{i^{-1}(\hat{\psi}, \hat{\lambda}_1, \dots, \hat{\lambda}_{18})\}}$, where i is the Fisher information matrix. For the conditional maximum likelihood estimator the standard error is obtained using $\sqrt{(-\partial^2 l_c(\psi)/\partial \psi^2)^{-1}}$, evaluated at $\hat{\psi}_c$. For the modified profile likelihood estimator, $\hat{\psi}_{mp}$, the standard error is obtained using $\sqrt{(\partial^2 l_{mp}(\psi)/\partial \psi^2)^{-1}}$, evaluated at $\hat{\psi}_{mp}$, i.e. using the Hessian. For the penalised (Firth) and penalised based on adjusted responses likelihood estimators, the standard errors are obtained using $\sqrt{\text{diag}\{i^{-1}(\hat{\psi}_*, \hat{\lambda}_{1,*}, \dots, \hat{\lambda}_{18,*})\}}$ and $\sqrt{\text{diag}\{i^{-1}(\hat{\psi}_a, \hat{\lambda}_{1,a}, \dots, \hat{\lambda}_{18,a})\}}$, respectively. Finally, for the indirect inference estimator the standard error is obtained using $\sqrt{(-\partial^2 l_c(\psi)/\partial \psi^2)^{-1}}$, evaluated at $\hat{\psi}_{a*}$. However according to Kuk (1995), to obtain the estimated standard error for the indirect inference estimator we need a further correction of the Fisher information using sandwich estimators of the variance based on the Godambe information because the second Bartlett identity no longer holds. Similarly, a Godambe information matrix would be better to use for the estimated standard error of the modified profile likelihood estimator. The estimated standard error reported in Lunardon (2018) for $\hat{\psi}_*$ is obtained using the Hessian matrix and is slightly different to our result.

To assess the reliability of the estimators we compute their actual bias and variance, conditioning on the observed totals as in (Lunardon 2018, Table 2), for a set of values of ψ ranging from -3 to 3. The conditional distribution of the sufficient statistic $T = \sum_{i=1}^{18} Y_{i1}$, which represents the total number of babies not crying in the experimental group, is the distribution of the sum of independent Bernoulli random variables with different probabilities. To calculate this distribution we use the R function *dkbinom* which gives the mass function of the sum of k independent Binomial random variables, with possibly different probabilities. This function implements the convolution algorithm of k binomials described in Butler and Stephens (2017).

The results reported in Table 2 of Lunardon (2018) are incorrect because Lunardon (2018) computes the conditional bias and variance of $\hat{\psi}_c$, $\hat{\psi}$, $\hat{\psi}_{mp}$ and $\hat{\psi}_*$ for $t \in \{1, \dots, 17\}$ but without rescaling and normalizing the conditional distribution of T . The corrected conditional summaries are reported in Table 3 with the addition of the penalised log-likelihood estimator based on adjusted responses, $\hat{\psi}_a$, and the indirect inference estimator, $\hat{\psi}_{a*}$, for various values of δ . We observe that for $\psi \in \{-2, -1, 0, 1, 2\}$ there exists a value of δ in the range $0.01 < \delta < 0.20$ such that $\hat{\psi}_a$ is less biased than any of $\hat{\psi}$, $\hat{\psi}_c$, $\hat{\psi}_{mp}$ and $\hat{\psi}_*$. In fact the effect of increasing the absolute value of ψ is to decrease the optimal δ value in terms of the bias of $\hat{\psi}_a$. For $\psi = 3$, the maximum likelihood estimator seems to be the least biased but this is due to the effect of infinite estimates at $t = 0$ and $t = 18$. In other words, removing these infinite estimates significantly lowers the average of the estimates for $t \in \{1, \dots, 17\}$. This is also true for the conditional log-likelihood estimator, $\hat{\psi}_c$.

Overall, for $\psi \in \{-2, -1, 0, 1\}$ the indirect inference estimator is an improvement over the penalised log-likelihood estimator based on adjusted responses in terms of bias for all values of δ considered. We notice that there exists a δ value in the range $0.01 < \delta < 0.20$ such that $\hat{\psi}_{a*}$ is less biased than $\hat{\psi}_a$. In fact the indirect inference estimator for $\psi \in \{-2, -1, 0, 1\}$ is competitive for all δ values considered, which makes it the best choice amongst estimators. However, for $\psi \in \{-3, 2, 3\}$ the best

choice of δ is the largest possible, in this case $\delta = 1$. It is worth noting that as in the complete enumeration study in Tables 5, 6, 7, 8, 9 and 10, it is also the case here that we observe that the bias and variance of the indirect inference estimator approaches that of the conditional log-likelihood estimator as δ increase to 1.

Table 4 reports the unconditional bias and variance of the modified profile, penalised (Firth) and penalised based on adjusted responses log-likelihood estimators for $t \in \{0, \dots, 18\}$. Even though $\hat{\psi}_*$ has smaller bias and variance than $\hat{\psi}_{mp}$ for all value of ψ , we again observe as in Table 3 that there exists a value of δ such that $\hat{\psi}_a$ is less biased than $\hat{\psi}_*$ for all values of ψ . We may conclude that $\hat{\psi}_a$ is preferable over $\hat{\psi}_*$ as its variance is also decreased at the optimal δ value for all values of ψ considered (except for $\psi = 0$). The relationship between the optimal δ value and the true value of ψ coincides with that in the conditional case (Table 3), i.e. the optimal δ , in terms of bias, decreases as the absolute value of ψ increases. Note that the indirect inference estimator has no solution at $t = 0$ and $t = 18$, and so it is excluded from the unconditional summaries in Table 4.

7 Conclusions

For the binomial matched pairs model, we evaluated the performance of a new penalised log-likelihood estimator of the log odds ratio which is based on an additive adjustment, $\delta > 0$, to the responses so as to avoid infinite estimates which is inherited by the maximum likelihood and conditional likelihood estimators. We calculated the probability limit of this estimator and showed that the maximum likelihood, conditional and modified profile log-likelihood estimators, when $m = 1$ for the latter, can be retrieved from this new estimator for certain values of δ . It was found that indirect inference estimation based on the new estimator is competitive for a wide range of values of δ .

It is worth investigating numerically whether, for a general value of m , there exists a δ that recovers the modified profile log-likelihood estimator of ψ from the penalised log-likelihood estimator based on adjusted responses, because this will imply that $\hat{\psi}_{mp}$ could be retrieved from $\hat{\psi}_a$ for any value of m not just in the special case of the binary matched pairs model where $m = 1$. One future direction is to investigate the performance of the estimators of the log odds ratio outside the setting of Lunardon (2018) for a general m_i and s_i . This is possible since in this general setting the distribution of the sufficient statistic $T = \sum_{i=1}^q y_{i1}$ can be obtained using the convolution method of Butler and Stephens (2017) and is Poisson binomial. Obtaining the probability limit of the indirect inference estimator in the setting of Lunardon (2018) is desired. The complete enumeration study in Sect. 5 may be expanded by considering other values of ψ , e.g. $\psi \in \{-3, -2, -1, 0, 1, 2, 3\}$. Finally, yet another possible future direction would be to investigate the performance of an alternative adjustment to the log-likelihood function where a small number $\delta > 0$ is added to each success but subtracted from each failure.

Table 2 Crying babies real data example: estimates of ψ and its standard error (in parentheses) derived from profile, conditional, modified profile, penalised (Firth), penalised based on adjusted responses log-likelihoods and indirect inference, respectively

δ	$\hat{\psi}$	$\hat{\psi}_c$	$\hat{\psi}_{mp}$	$\hat{\psi}_*$	$\hat{\psi}_a$	$\hat{\psi}_{a*}$
0.05	1.432 (0.734)	1.256 (0.686)	1.270 (0.689)	1.156 (0.666)	1.174 (0.656)	1.157 (0.669)
0.10					0.986 (0.602)	1.184 (0.674)
0.15					0.841 (0.561)	1.201 (0.676)
0.20					0.726 (0.529)	1.212 (0.678)
0.25					0.633 (0.502)	1.220 (0.680)
0.30					0.555 (0.480)	1.226 (0.681)
0.35					0.490 (0.461)	1.230 (0.682)
0.40					0.435 (0.445)	1.234 (0.682)
0.45					0.387 (0.430)	1.237 (0.683)
0.50					0.345 (0.417)	1.239 (0.683)
0.55					0.309 (0.405)	1.241 (0.684)
0.60					0.277 (0.395)	1.243 (0.684)
0.65					0.249 (0.385)	1.245 (0.684)
0.70					0.224 (0.376)	1.246 (0.684)
0.75					0.201 (0.368)	1.247 (0.684)
0.80					0.181 (0.361)	1.248 (0.685)
0.85					0.162 (0.354)	1.248 (0.685)
0.90					0.146 (0.347)	1.249 (0.685)
0.95					0.131 (0.341)	1.250 (0.685)
1.00					0.117 (0.335)	1.251 (0.685)

Table 3 Crying babies real data example: conditional bias and variance (in parentheses) of estimators as the true log odds ratio ψ varies; estimators derived from profile, conditional, modified profile, penalised (Firth), penalised based on adjusted responses log-likelihoods and indirect inference are denoted by $\hat{\psi}$, $\hat{\psi}_c$, $\hat{\psi}_{mp}$, $\hat{\psi}_*$, $\hat{\psi}_a$, $\tilde{\psi}_{a*}$, respectively, and all entries are multiplied by 10. For $\hat{\psi}_a$ and $\tilde{\psi}_{a*}$, the smallest bias value in each column is given in bold face

	ψ						
	-3	-2	-1	0	1	2	3
$\hat{\psi}$	-8.97 (9.56)	-5.59 (7.60)	-2.02 (4.48)	0.37 (4.40)	2.47 (6.25)	2.30 (5.71)	-2.53 (2.98)
$\hat{\psi}_c$	-0.05 (4.18)	-0.69 (4.01)	-0.12 (3.00)	0.36 (3.30)	0.88 (4.65)	-0.71 (4.06)	-6.36 (2.06)
$\hat{\psi}_{mp}$	-0.79 (3.89)	-1.39 (4.22)	-0.42 (3.23)	0.34 (3.42)	1.00 (4.75)	-0.51 (4.13)	-6.12 (2.10)
$\hat{\psi}_*$	0.56 (3.35)	-0.62 (3.77)	-0.17 (2.94)	0.07 (3.02)	-0.12 (3.75)	-2.71 (2.95)	-9.04 (1.42)
$\hat{\psi}_a$	-7.51 (7.96)	-4.99 (6.83)	-1.88 (4.27)	0.22 (4.17)	1.86 (5.56)	1.02 (4.77)	-4.27 (2.41)
δ	0.01						-5.77 (2.00)
	0.02	-6.23 (6.77)	-1.75 (4.07)	0.08 (3.95)	1.31 (4.99)	-0.10 (4.06)	-7.06 (1.69)
	0.03	-5.09 (5.86)	-3.94 (5.68)	-0.06 (3.75)	0.80 (4.53)	-1.08 (3.52)	-8.20 (1.46)
	0.04	-4.07 (5.14)	-3.46 (5.23)	-0.18 (3.57)	0.33 (4.15)	-1.95 (3.10)	-9.22 (1.27)
	0.05	-3.13 (4.56)	-3.01 (4.84)	-0.30 (3.41)	-0.10 (3.82)	-2.74 (2.75)	-10.14 (1.12)
	0.06	-2.28 (4.08)	-2.59 (4.51)	-0.41 (3.26)	-0.50 (3.54)	-3.47 (2.47)	-10.97 (1.00)
	0.07	-1.49 (3.68)	-2.20 (4.21)	-0.51 (3.12)	-0.87 (3.29)	-4.13 (2.24)	-11.73 (0.90)
	0.08	-0.76 (3.35)	-1.82 (3.94)	-0.61 (2.99)	-1.22 (3.07)	-4.73 (2.04)	-12.43 (0.82)
	0.09	-0.08 (3.06)	-1.46 (3.71)	-0.71 (2.86)	-1.55 (2.88)	-5.30 (1.87)	-13.07 (0.75)
	0.10	0.55 (2.81)	-1.12 (3.50)	-0.80 (2.75)	-1.86 (2.71)	-5.82 (1.72)	-15.70 (0.50)
	0.15	3.22 (1.96)	0.37 (2.69)	-1.18 (2.28)	-3.18 (2.07)	-7.99 (1.21)	-17.65 (0.37)
	0.20	5.28 (1.46)	1.59 (2.16)	0.03 (2.06)	-1.48 (1.92)	-9.63 (0.92)	-26.96 (0.05)
	1.00	17.07 (0.19)	9.53 (0.37)	-2.73 (0.41)	-9.52 (0.29)	-17.70 (0.13)	

Table 3 continued

$\tilde{\psi}_{a*}$	δ	ψ						
		-3	-2	-1	0	1	2	3
	0.01	2.72 (2.60)	0.51 (3.00)	0.13 (2.64)	-0.01 (2.83)	-0.62 (3.30)	-3.75 (2.46)	-10.41 (1.17)
	0.02	2.26 (2.93)	0.37 (3.17)	0.11 (2.67)	0.02 (2.88)	-0.40 (3.57)	-3.16 (2.83)	-9.55 (1.39)
	0.03	1.97 (3.14)	0.28 (3.28)	0.10 (2.69)	0.04 (2.92)	-0.25 (3.74)	-2.80 (3.06)	-9.04 (1.52)
	0.04	1.75 (3.29)	0.20 (3.36)	0.09 (2.71)	0.07 (2.95)	-0.14 (3.87)	-2.54 (3.22)	-8.68 (1.61)
	0.05	1.57 (3.41)	0.14 (3.43)	0.08 (2.72)	0.08 (2.97)	-0.05 (3.96)	-2.34 (3.34)	-8.40 (1.68)
	0.06	1.43 (3.50)	0.08 (3.48)	0.07 (2.74)	0.10 (2.99)	0.02 (4.03)	-2.19 (3.43)	-8.19 (1.73)
	0.07	1.31 (3.57)	0.03 (3.53)	0.06 (2.75)	0.11 (3.01)	0.08 (4.09)	-2.05 (3.51)	-8.02 (1.78)
	0.08	1.21 (3.63)	-0.01 (3.57)	0.06 (2.76)	0.13 (3.03)	0.14 (4.15)	-1.94 (3.57)	-7.87 (1.81)
	0.09	1.12 (3.69)	-0.04 (3.60)	0.05 (2.77)	0.14 (3.04)	0.18 (4.19)	-1.85 (3.62)	-7.75 (1.84)
	0.10	1.05 (3.73)	-0.08 (3.63)	0.04 (2.78)	0.15 (3.06)	0.23 (4.23)	-1.77 (3.66)	-7.64 (1.86)
	0.15	0.77 (3.88)	-0.21 (3.74)	0.02 (2.83)	0.19 (3.11)	0.38 (4.35)	-1.48 (3.80)	-7.27 (1.93)
	0.20	0.59 (3.96)	-0.29 (3.80)	0.00 (2.86)	0.23 (3.14)	0.48 (4.43)	-1.30 (3.87)	-7.06 (1.97)
	1.00	0.05 (4.16)	-0.62 (3.99)	-0.09 (2.98)	0.35 (3.28)	0.83 (4.63)	-0.78 (4.04)	-6.44 (2.06)

Table 4 Crying babies real data example: unconditional bias and variance (in parentheses) of estimators as the true log odds ratio ψ varies; estimators derived from modified profile, penalised (Firth) and penalised based on adjusted responses log-likelihoods are denoted by $\hat{\psi}_{mp}$, $\hat{\psi}_*$, $\hat{\psi}_a$, respectively, and all entries are multiplied by 10. For $\hat{\psi}_a$, the smallest bias value in each column is given in bold face

		ψ						
		-3	-2	-1	0	1	2	3
$\hat{\psi}_{mp}$		-2.65 (6.86)	-1.50 (4.52)	-0.42 (3.23)	0.35 (3.44)	1.46 (6.60)	4.26 (17.04)	7.10 (21.60)
$\hat{\psi}_*$		-1.13 (5.82)	-0.72 (4.03)	-0.17 (2.95)	0.07 (3.03)	0.18 (4.55)	0.18 (7.49)	-1.48 (7.45)
$\hat{\psi}_a$	δ	-7.87 (12.84)	-4.09 (6.29)	-1.62 (3.89)	-0.05 (3.77)	1.13 (5.47)	2.07 (8.88)	1.13 (8.79)
		-6.46 (10.22)	-3.60 (5.71)	-1.49 (3.71)	-0.18 (3.58)	0.63 (4.87)	0.75 (6.98)	-1.30 (6.37)
		-5.83 (9.26)	-3.36 (5.46)	-1.43 (3.63)	-0.24 (3.50)	0.40 (4.63)	0.17 (6.30)	-2.31 (5.56)
		-5.24 (8.46)	-3.14 (5.23)	-1.37 (3.56)	-0.30 (3.42)	0.17 (4.41)	-0.36 (5.74)	-3.22 (4.91)
		-4.18 (7.19)	-2.71 (4.83)	-1.26 (3.41)	-0.41 (3.26)	-0.25 (4.03)	-1.32 (4.86)	-4.80 (3.95)
		-0.82 (4.39)	-1.21 (3.68)	-0.84 (2.90)	-0.79 (2.75)	-1.67 (3.00)	-4.24 (2.96)	-9.28 (2.08)
		2.17 (2.84)	0.30 (2.81)	-0.37 (2.43)	-1.18 (2.28)	-3.03 (2.25)	-6.77 (1.93)	-12.85 (1.23)
		4.42 (2.04)	1.53 (2.24)	0.03 (2.06)	-1.48 (1.93)	-4.08 (1.78)	-8.62 (1.39)	-15.33 (0.83)

Appendix A: Results tables for complete enumeration study

See Tables 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17.

Table 5 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the bias and variance (in parentheses) of estimators derived from profile, conditional, modified profile, penalised (Firth) and penalised based on adjusted responses log-likelihoods, with all entries multiplied by 10; seventh column show the bias and variance (in parentheses) of the indirect inference $\hat{\psi}_{a*}$ using the conditional model, with all entries multiplied by 10

	$\hat{\psi}$	$\hat{\psi}_{\mathbf{c}}$	$\hat{\psi}_{\mathbf{mp}}$	$\hat{\psi}_{*}$	$\hat{\psi}_{\mathbf{a}}$	$\hat{\psi}_{\mathbf{a}*}$
δ	$q = 30, m = 1$					
0.05	10.89(7.87)	0.44(1.97)	2.91(2.33)	2.76(2.27)	8.49(5.58)	0.08(1.80)
0.10					6.65(4.25)	0.17(1.85)
0.15					5.16(3.39)	0.22(1.88)
0.20					3.94(2.78)	0.26(1.90)
0.25					2.91(2.33)	0.29(1.91)
0.30					2.03(1.99)	0.31(1.92)
0.35					1.27(1.72)	0.33(1.92)
0.40					0.60(1.50)	0.34(1.93)
0.45					0.01(1.33)	0.35(1.93)
0.50					-0.52(1.18)	0.36(1.94)
0.55					-0.99(1.06)	0.37(1.94)
0.60					-1.42(0.95)	0.38(1.94)
0.65					-1.81(0.87)	0.38(1.95)
0.70					-2.16(0.79)	0.39(1.95)
0.75					-2.49(0.72)	0.39(1.95)
0.80					-2.79(0.66)	0.40(1.95)
0.85					-3.06(0.61)	0.40(1.95)
0.90					-3.32(0.57)	0.40(1.95)
0.95					-3.55(0.53)	0.41(1.95)
1.00					-3.77(0.49)	0.41(1.96)
δ	$q = 30, m = 3$					
0.05	3.56(3.06)	0.44(1.97)	0.69(1.82)	0.48(1.72)	2.26(2.30)	0.15(1.82)
0.10					1.24(1.84)	0.22(1.87)
0.15					0.41(1.53)	0.27(1.89)
0.20					-0.29(1.30)	0.30(1.91)
0.25					-0.88(1.13)	0.32(1.92)
0.30					-1.40(0.99)	0.34(1.93)
0.35					-1.85(0.88)	0.35(1.93)
0.40					-2.25(0.79)	0.37(1.94)
0.45					-2.61(0.71)	0.37(1.94)
0.50					-2.94(0.64)	0.38(1.95)

Table 5 continued

	$\hat{\psi}$	$\hat{\psi}_{\mathbf{c}}$	$\hat{\psi}_{\mathbf{mp}}$	$\hat{\psi}_{*}$	$\hat{\psi}_{\mathbf{a}}$	$\tilde{\psi}_{\mathbf{a}*}$
0.55					− 3.23 (0.59)	0.39 (1.95)
0.60					− 3.50 (0.54)	0.39 (1.95)
0.65					− 3.74 (0.50)	0.40 (1.95)
0.70					− 3.97 (0.46)	0.40 (1.95)
0.75					− 4.18 (0.43)	0.41 (1.95)
0.80					− 4.37 (0.40)	0.41 (1.95)
0.85					− 4.56 (0.37)	0.41 (1.96)
0.90					− 4.72 (0.35)	0.41 (1.96)
0.95					− 4.88 (0.33)	0.42 (1.96)
1.00					− 5.03 (0.31)	0.42 (1.96)

Table 6 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the bias and variance (in parentheses) of estimators derived from profile, conditional, modified profile, penalised (Firth) and penalised based on adjusted responses log-likelihoods, with all entries multiplied by 10; seventh column show the bias and variance (in parentheses) of the indirect inference $\hat{\psi}_{\mathbf{a}*}$ using the conditional model, with all entries multiplied by 10

	$\hat{\psi}$	$\hat{\psi}_{\mathbf{c}}$	$\hat{\psi}_{\mathbf{mp}}$	$\hat{\psi}_{*}$	$\hat{\psi}_{\mathbf{a}}$	$\tilde{\psi}_{\mathbf{a}*}$
δ	$q = 30, m = 11$					
0.05	1.28 (2.23)	0.44 (1.97)	0.46 (1.94)	0.09 (1.74)	0.08 (1.62)	0.11 (1.81)
0.10					− 0.85 (1.26)	0.19 (1.85)
0.15					− 1.60 (1.02)	0.24 (1.88)
0.20					− 2.22 (0.85)	0.27 (1.90)
0.25					− 2.74 (0.73)	0.30 (1.91)
0.30					− 3.19 (0.63)	0.32 (1.92)
0.35					− 3.58 (0.55)	0.34 (1.93)
0.40					− 3.92 (0.49)	0.35 (1.93)
0.45					− 4.22 (0.44)	0.36 (1.94)
0.50					− 4.49 (0.40)	0.37 (1.94)
0.55					− 4.73 (0.36)	0.38 (1.94)
0.60					− 4.95 (0.33)	0.39 (1.95)
0.65					− 5.15 (0.30)	0.39 (1.95)
0.70					− 5.34 (0.28)	0.40 (1.95)
0.75					− 5.50 (0.26)	0.40 (1.95)
0.80					− 5.66 (0.24)	0.40 (1.95)
0.85					− 5.80 (0.22)	0.41 (1.95)
0.90					− 5.94 (0.21)	0.41 (1.96)
0.95					− 6.06 (0.20)	0.41 (1.96)
1.00					− 6.17 (0.18)	0.42 (1.96)

Table 6 continued

	$\hat{\psi}$	$\hat{\psi}_{\mathbf{c}}$	$\hat{\psi}_{\mathbf{mp}}$	$\hat{\psi}_{*}$	$\hat{\psi}_{\mathbf{a}}$	$\tilde{\psi}_{\mathbf{a}*}$
δ	$q = 30, m = 39$					
0.05	0.68 (2.04)	0.44 (1.97)	0.44 (1.96)	0.03 (1.74)	-0.52 (1.46)	0.09 (1.80)
0.10					-1.44 (1.12)	0.17 (1.85)
0.15					-2.18 (0.90)	0.23 (1.88)
0.20					-2.79 (0.74)	0.26 (1.90)
0.25					-3.31 (0.62)	0.29 (1.91)
0.30					-3.75 (0.53)	0.31 (1.92)
0.35					-4.13 (0.46)	0.33 (1.93)
0.40					-4.47 (0.41)	0.34 (1.93)
0.45					-4.77 (0.36)	0.36 (1.94)
0.50					-5.03 (0.32)	0.37 (1.94)
0.55					-5.27 (0.29)	0.37 (1.94)
0.60					-5.48 (0.26)	0.38 (1.95)
0.65					-5.68 (0.24)	0.39 (1.95)
0.70					-5.85 (0.22)	0.39 (1.95)
0.75					-6.02 (0.20)	0.40 (1.95)
0.80					-6.17 (0.19)	0.40 (1.95)
0.85					-6.30 (0.17)	0.40 (1.95)
0.90					-6.43 (0.16)	0.41 (1.95)
0.95					-6.55 (0.15)	0.41 (1.96)
1.00					-6.66 (0.14)	0.41 (1.96)

Table 7 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the bias and variance (in parentheses) of estimators derived from profile, conditional, modified profile, penalised (Firth) and penalised based on adjusted responses log-likelihoods, with all entries multiplied by 10; seventh column show the bias and variance (in parentheses) of the indirect inference $\tilde{\psi}_{\mathbf{a}*}$ using the conditional model, with all entries multiplied by 10

	$\hat{\psi}$	$\hat{\psi}_{\mathbf{c}}$	$\hat{\psi}_{\mathbf{mp}}$	$\hat{\psi}_{*}$	$\hat{\psi}_{\mathbf{a}}$	$\tilde{\psi}_{\mathbf{a}*}$
δ	$q = 100, m = 1$					
0.05	10.24 (2.11)	0.12 (0.53)	2.79 (0.69)	2.74 (0.68)	8.08 (1.57)	0.02 (0.52)
0.10					6.36 (1.23)	0.04 (0.52)
0.15					4.96 (0.99)	0.06 (0.52)
0.20					3.78 (0.82)	0.07 (0.52)
0.25					2.79 (0.69)	0.08 (0.52)
0.30					1.93 (0.59)	0.08 (0.52)

Table 7 continued

	$\hat{\psi}$	$\hat{\psi}_c$	$\hat{\psi}_{mp}$	$\hat{\psi}_*$	$\hat{\psi}_a$	$\tilde{\psi}_{a*}$
0.35					1.19 (0.51)	0.09 (0.52)
0.40					0.53 (0.45)	0.09 (0.53)
0.45					-0.05 (0.40)	0.09 (0.53)
0.50					-0.57 (0.35)	0.10 (0.53)
0.55					-1.03 (0.32)	0.10 (0.53)
0.60					-1.46 (0.29)	0.10 (0.53)
0.65					-1.84 (0.26)	0.10 (0.53)
0.70					-2.19 (0.24)	0.10 (0.53)
0.75					-2.51 (0.22)	0.11 (0.53)
0.80					-2.81 (0.20)	0.11 (0.53)
0.85					-3.08 (0.18)	0.11 (0.53)
0.90					-3.33 (0.17)	0.11 (0.53)
0.95					-3.56 (0.16)	0.11 (0.53)
1.00					-3.78 (0.15)	0.11 (0.53)
δ	$q = 100, m = 3$					
0.05	3.23 (0.84)	0.12 (0.53)	0.48 (0.51)	0.42 (0.50)	2.05 (0.66)	0.04 (0.52)
0.10					1.09 (0.54)	0.06 (0.52)
0.15					0.30 (0.45)	0.07 (0.52)
0.20					-0.37 (0.38)	0.08 (0.52)
0.25					-0.95 (0.33)	0.09 (0.52)
0.30					-1.45 (0.29)	0.09 (0.53)
0.35					-1.89 (0.26)	0.10 (0.53)
0.40					-2.29 (0.23)	0.10 (0.53)
0.45					-2.64 (0.21)	0.10 (0.53)
0.50					-2.96 (0.19)	0.10 (0.53)
0.55					-3.25 (0.18)	0.10 (0.53)
0.60					-3.52 (0.16)	0.11 (0.53)
0.65					-3.76 (0.15)	0.11 (0.53)
0.70					-3.99 (0.14)	0.11 (0.53)
0.75					-4.19 (0.13)	0.11 (0.53)
0.80					-4.39 (0.12)	0.11 (0.53)
0.85					-4.57 (0.11)	0.11 (0.53)
0.90					-4.73 (0.10)	0.11 (0.53)
0.95					-4.89 (0.10)	0.11 (0.53)
1.00					-5.04 (0.09)	0.11 (0.53)

Table 8 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the bias and variance (in parentheses) of estimators derived from profile, conditional, modified profile, penalised (Firth) and penalised based on adjusted responses log-likelihoods, with all entries multiplied by 10; seventh column show the bias and variance (in parentheses) of the indirect inference $\hat{\psi}_{a*}$ using the conditional model, with all entries multiplied by 10

	$\hat{\psi}$	$\hat{\psi}_c$	$\hat{\psi}_{mp}$	$\hat{\psi}_*$	$\hat{\psi}_a$	$\hat{\psi}_{a*}$
δ	$q = 100, m = 11$					
0.05	0.96(0.61)	0.12(0.53)	0.15(0.52)	0.05(0.51)	-0.13(0.46)	0.03(0.52)
0.10					-0.99(0.36)	0.05(0.52)
0.15					-1.70(0.30)	0.06(0.52)
0.20					-2.30(0.25)	0.07(0.52)
0.25					-2.80(0.21)	0.08(0.52)
0.30					-3.24(0.19)	0.09(0.52)
0.35					-3.62(0.16)	0.09(0.53)
0.40					-3.95(0.15)	0.09(0.53)
0.45					-4.25(0.13)	0.10(0.53)
0.50					-4.51(0.12)	0.10(0.53)
0.55					-4.75(0.11)	0.10(0.53)
0.60					-4.97(0.10)	0.10(0.53)
0.65					-5.17(0.09)	0.11(0.53)
0.70					-5.35(0.08)	0.11(0.53)
0.75					-5.52(0.08)	0.11(0.53)
0.80					-5.67(0.07)	0.11(0.53)
0.85					-5.81(0.07)	0.11(0.53)
0.90					-5.94(0.06)	0.11(0.53)
0.95					-6.07(0.06)	0.11(0.53)
1.00					-6.18(0.06)	0.11(0.53)
δ	$q = 100, m = 39$					
0.05	0.36(0.55)	0.12(0.53)	0.12(0.53)	0.01(0.51)	-0.72(0.41)	0.03(0.52)
0.10					-1.58(0.32)	0.05(0.52)
0.15					-2.29(0.26)	0.06(0.52)
0.20					-2.87(0.22)	0.07(0.52)
0.25					-3.37(0.18)	0.08(0.52)
0.30					-3.80(0.16)	0.08(0.52)
0.35					-4.17(0.14)	0.09(0.53)
0.40					-4.50(0.12)	0.09(0.53)
0.45					-4.79(0.11)	0.10(0.53)
0.50					-5.05(0.10)	0.10(0.53)
0.55					-5.29(0.09)	0.10(0.53)
0.60					-5.50(0.08)	0.10(0.53)
0.65					-5.69(0.07)	0.10(0.53)
0.70					-5.87(0.07)	0.11(0.53)
0.75					-6.03(0.06)	0.11(0.53)

Table 8 continued

	$\hat{\psi}$	$\hat{\psi}_{\mathbf{c}}$	$\hat{\psi}_{\mathbf{mp}}$	$\hat{\psi}_{*}$	$\hat{\psi}_{\mathbf{a}}$	$\tilde{\psi}_{\mathbf{a}*}$
0.80					-6.18 (0.06)	0.11 (0.53)
0.85					-6.31 (0.05)	0.11 (0.53)
0.90					-6.44 (0.05)	0.11 (0.53)
0.95					-6.56 (0.05)	0.11 (0.53)
1.00					-6.67 (0.04)	0.11 (0.53)

Table 9 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the bias and variance (in parentheses) of estimators derived from profile, conditional, modified profile, penalised (Firth) and penalised based on adjusted responses log-likelihoods, with all entries multiplied by 10; seventh column show the bias and variance (in parentheses) of the indirect inference $\hat{\psi}_{a*}$ using the conditional model, with all entries multiplied by 10

	$\hat{\psi}$	$\hat{\psi}_{\mathbf{c}}$	$\hat{\psi}_{\mathbf{mp}}$	$\hat{\psi}_{*}$	$\hat{\psi}_{\mathbf{a}}$	$\tilde{\psi}_{\mathbf{a}*}$
δ	$q = 1000, m = 1$					
0.05	10.02 (0.20)	0.01 (0.05)	2.74 (0.07)	2.74 (0.07)	7.93 (0.15)	0.00 (0.05)
0.10					6.25 (0.12)	0.00 (0.05)
0.15					4.88 (0.10)	0.01 (0.05)
0.20					3.72 (0.08)	0.01 (0.05)
0.25					2.74 (0.07)	0.01 (0.05)
0.30					1.90 (0.06)	0.01 (0.05)
0.35					1.16 (0.05)	0.01 (0.05)
0.40					0.51 (0.04)	0.01 (0.05)
0.45					-0.07 (0.04)	0.01 (0.05)
0.50					-0.59 (0.04)	0.01 (0.05)
0.55					-1.05 (0.03)	0.01 (0.05)
0.60					-1.47 (0.03)	0.01 (0.05)
0.65					-1.85 (0.03)	0.01 (0.05)
0.70					-2.20 (0.02)	0.01 (0.05)
0.75					-2.52 (0.02)	0.01 (0.05)
0.80					-2.81 (0.02)	0.01 (0.05)
0.85					-3.08 (0.02)	0.01 (0.05)
0.90					-3.34 (0.02)	0.01 (0.05)
0.95					-3.57 (0.02)	0.01 (0.05)
1.00					-3.79 (0.01)	0.01 (0.05)
δ	$q = 1000, m = 3$					
0.05	3.12 (0.08)	0.01 (0.05)	0.40 (0.05)	0.40 (0.05)	1.97 (0.06)	0.00 (0.05)
0.10					1.04 (0.05)	0.01 (0.05)
0.15					0.26 (0.04)	0.01 (0.05)
0.20					-0.40 (0.04)	0.01 (0.05)
0.25					-0.97 (0.03)	0.01 (0.05)

Table 9 continued

	$\hat{\psi}$	$\hat{\psi}_{\mathbf{c}}$	$\hat{\psi}_{\mathbf{mp}}$	$\hat{\psi}_{*}$	$\hat{\psi}_{\mathbf{a}}$	$\tilde{\psi}_{\mathbf{a}*}$
0.30					− 1.47 (0.03)	0.01 (0.05)
0.35					− 1.91 (0.03)	0.01 (0.05)
0.40					− 2.30 (0.02)	0.01 (0.05)
0.45					− 2.65 (0.02)	0.01 (0.05)
0.50					− 2.97 (0.02)	0.01 (0.05)
0.55					− 3.26 (0.02)	0.01 (0.05)
0.60					− 3.53 (0.02)	0.01 (0.05)
0.65					− 3.77 (0.01)	0.01 (0.05)
0.70					− 3.99 (0.01)	0.01 (0.05)
0.75					− 4.20 (0.01)	0.01 (0.05)
0.80					− 4.39 (0.01)	0.01 (0.05)
0.85					− 4.57 (0.01)	0.01 (0.05)
0.90					− 4.74 (0.01)	0.01 (0.05)
0.95					− 4.89 (0.01)	0.01 (0.05)
1.00					− 5.04 (0.01)	0.01 (0.05)

Table 10 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the bias and variance (in parentheses) of estimators derived from profile, conditional, modified profile, penalised (Firth) and penalised based on adjusted responses log-likelihoods, with all entries multiplied by 10; seventh column show the bias and variance (in parentheses) of the indirect inference $\tilde{\psi}_{\mathbf{a}*}$ using the conditional model, with all entries multiplied by 10

	$\hat{\psi}$	$\hat{\psi}_{\mathbf{c}}$	$\hat{\psi}_{\mathbf{mp}}$	$\hat{\psi}_{*}$	$\hat{\psi}_{\mathbf{a}}$	$\tilde{\psi}_{\mathbf{a}*}$
δ	$q = 1000, m = 11$					
0.05	0.85 (0.06)	0.01 (0.05)	0.04 (0.05)	0.03 (0.05)	− 0.20 (0.04)	0.00 (0.05)
0.10					− 1.05 (0.04)	0.00 (0.05)
0.15					− 1.74 (0.03)	0.01 (0.05)
0.20					− 2.33 (0.02)	0.01 (0.05)
0.25					− 2.82 (0.02)	0.01 (0.05)
0.30					− 3.26 (0.02)	0.01 (0.05)
0.35					− 3.63 (0.02)	0.01 (0.05)
0.40					− 3.96 (0.01)	0.01 (0.05)
0.45					− 4.26 (0.01)	0.01 (0.05)
0.50					− 4.52 (0.01)	0.01 (0.05)
0.55					− 4.76 (0.01)	0.01 (0.05)
0.60					− 4.98 (0.01)	0.01 (0.05)
0.65					− 5.17 (0.01)	0.01 (0.05)
0.70					− 5.35 (0.01)	0.01 (0.05)
0.75					− 5.52 (0.01)	0.01 (0.05)

Table 10 continued

	$\hat{\psi}$	$\hat{\psi}_c$	$\hat{\psi}_{mp}$	$\hat{\psi}_*$	$\hat{\psi}_a$	$\tilde{\psi}_{a*}$
0.80					-5.67 (0.01)	0.01 (0.05)
0.85					-5.81 (0.01)	0.01 (0.05)
0.90					-5.95 (0.01)	0.01 (0.05)
0.95					-6.07 (0.01)	0.01 (0.05)
1.00					-6.18 (0.01)	0.01 (0.05)
δ	$q = 1000, m = 39$					
0.05	0.25 (0.05)	0.01 (0.05)	0.01 (0.05)	0.00 (0.05)	-0.80 (0.04)	0.00 (0.05)
0.10					-1.64 (0.03)	0.00 (0.05)
0.15					-2.33 (0.03)	0.01 (0.05)
0.20					-2.90 (0.02)	0.01 (0.05)
0.25					-3.39 (0.02)	0.01 (0.05)
0.30					-3.82 (0.02)	0.01 (0.05)
0.35					-4.19 (0.01)	0.01 (0.05)
0.40					-4.51 (0.01)	0.01 (0.05)
0.45					-4.80 (0.01)	0.01 (0.05)
0.50					-5.06 (0.01)	0.01 (0.05)
0.55					-5.29 (0.01)	0.01 (0.05)
0.60					-5.50 (0.01)	0.01 (0.05)
0.65					-5.70 (0.01)	0.01 (0.05)
0.70					-5.87 (0.01)	0.01 (0.05)
0.75					-6.03 (0.01)	0.01 (0.05)
0.80					-6.18 (0.01)	0.01 (0.05)
0.85					-6.32 (0.01)	0.01 (0.05)
0.90					-6.44 (0.00)	0.01 (0.05)
0.95					-6.56 (0.00)	0.01 (0.05)
1.00					-6.67 (0.00)	0.01 (0.05)

Table 11 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the coverage probability and average length (in parentheses) of confidence intervals with nominal level 95% derived from profile, conditional, modified profile, penalized (Firth) and penalised based on adjusted responses log-likelihood ratios, with all coverage probabilities multiplied by 100

	W	W_c	W_{mp}	W_*	W_a
δ	$q = 30, m = 1$				
0.31	57.8 (2.4)	96.2 (1.7)	85.0 (1.8)	93.0 (1.7)	92.0 (1.7)
0.32					92.0 (1.7)
0.33					92.0 (1.7)
0.34					92.0 (1.6)
0.35					92.0 (1.6)
0.36					96.2 (1.6)

Table 11 continued

	W	W_c	W_{mp}	W_*	W_a
0.37					96.2(1.6)
0.38					96.2(1.6)
0.39					96.2(1.6)
0.40					96.2(1.6)
0.41					96.2(1.6)
0.42					96.2(1.6)
0.43					96.2(1.5)
0.44					96.2(1.5)
0.45					96.2(1.5)
0.46					97.8(1.5)
0.47					97.8(1.5)
0.48					97.8(1.5)
0.49					95.6(1.5)
0.50					95.6(1.5)
δ	$q = 30, m = 3$				
0.06	85.0(1.9)	96.2(1.7)	96.2(1.7)	96.2(1.6)	92.0(1.7)
0.07					92.0(1.7)
0.08					92.0(1.7)
0.09					92.0(1.7)
0.10					92.0(1.7)
0.11					96.2(1.6)
0.12					96.2(1.6)
0.13					96.2(1.6)
0.14					96.2(1.6)
0.15					96.2(1.6)
0.16					96.2(1.6)
0.17					96.2(1.6)
0.18					96.2(1.5)
0.19					97.8(1.5)
0.20					95.6(1.5)
0.21					95.6(1.5)
0.22					95.6(1.5)
0.23					95.6(1.5)
0.24					95.6(1.5)
0.25					95.6(1.5)

Table 12 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the coverage probability and average length (in parentheses) of confidence intervals with nominal level 95% derived from profile, conditional, modified profile, penalized (Firth) and penalised based on adjusted responses log-likelihood ratios, with all coverage probabilities multiplied by 100

	W	W_c	W_{mp}	W_*	W_a
δ	$q = 30, m = 11$				
0.01	92.0(1.8)	96.2(1.7)	96.2(1.7)	93.9(1.7)	96.2(1.7)
0.02					96.2(1.7)
0.03					96.2(1.7)
0.04					96.2(1.6)
0.05					96.2(1.6)
0.06					93.9(1.6)
0.07					93.9(1.6)
0.08					95.6(1.6)
0.09					95.6(1.5)
0.10					95.6(1.5)
0.11					95.6(1.5)
0.12					95.6(1.5)
0.13					95.6(1.5)
0.14					96.1(1.5)
0.15					96.1(1.4)
0.16					96.1(1.4)
0.17					96.1(1.4)
0.18					96.1(1.4)
0.19					91.7(1.4)
0.20					91.8(1.4)
δ	$q = 30, m = 39$				
0.01	96.2(1.7)	96.2(1.7)	96.2(1.7)	93.9(1.6)	96.2(1.7)
0.02					93.9(1.7)
0.03					93.9(1.6)
0.04					93.9(1.6)
0.05					95.6(1.6)
0.06					95.6(1.6)
0.07					95.6(1.5)
0.08					95.6(1.5)
0.09					95.6(1.5)
0.10					95.6(1.5)
0.11					96.1(1.5)
0.12					96.1(1.4)
0.13					96.1(1.4)

Table 12 continued

	W	W_c	W_{mp}	W_*	W_a
0.14					91.7 (1.4)
0.14					91.7 (1.4)
0.16					91.8 (1.4)
0.17					91.8 (1.4)
0.18					91.8 (1.4)
0.19					91.8 (1.3)
0.20					91.8 (1.3)

Table 13 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the coverage probability and average length (in parentheses) of confidence intervals with nominal level 95% derived from profile, conditional, modified profile, penalized (Firth) and penalised based on adjusted responses log-likelihood ratios, with all coverage probabilities multiplied by 100

	W	W_c	W_{mp}	W_*	W_a
δ	$q = 100, m = 1$				
0.31	15.0 (1.3)	94.6 (0.9)	77.3 (1.0)	77.3 (1.0)	88.4 (0.9)
0.32					88.4 (0.9)
0.33					88.4 (0.9)
0.34					92.3 (0.9)
0.35					92.3 (0.9)
0.36					91.8 (0.9)
0.37					94.6 (0.9)
0.38					94.6 (0.9)
0.39					94.6 (0.9)
0.40					94.6 (0.9)
0.41					95.7 (0.9)
0.42					95.7 (0.8)
0.43					95.7 (0.8)
0.44					96.9 (0.8)
0.45					96.9 (0.8)
0.46					95.7 (0.8)
0.47					96.4 (0.8)
0.48					96.4 (0.8)
0.49					96.4 (0.8)
0.50					96.4 (0.8)
δ	$q = 100, m = 3$				
0.06	69.9 (1.0)	94.6 (0.9)	93.9 (0.9)	95.7 (0.9)	88.4 (0.9)
0.07					88.4 (0.9)

Table 13 continued

	W	W_c	W_{mp}	W_*	W_a
0.08					88.4 (0.9)
0.09					91.8 (0.9)
0.10					91.8 (0.9)
0.11					94.6 (0.9)
0.12					94.6 (0.9)
0.13					94.6 (0.9)
0.14					95.7 (0.9)
0.15					95.7 (0.9)
0.16					95.7 (0.9)
0.17					96.9 (0.8)
0.18					95.7 (0.8)
0.19					95.7 (0.8)
0.20					96.4 (0.8)
0.21					96.4 (0.8)
0.22					94.7 (0.8)
0.23					95.1 (0.8)
0.24					95.1 (0.8)
0.25					95.1 (0.8)

Table 14 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the coverage probability and average length (in parentheses) of confidence intervals with nominal level 95% derived from profile, conditional, modified profile, penalized (Firth) and penalised based on adjusted responses log-likelihood ratios, with all coverage probabilities multiplied by 100

	W	W_c	W_{mp}	W_*	W_a
δ	$q = 100, m = 11$				
0.01	91.8 (0.9)	94.6 (0.9)	94.6 (0.9)	94.6 (0.9)	93.9 (0.9)
0.02					93.9 (0.9)
0.03					95.7 (0.9)
0.04					94.6 (0.9)
0.05					95.7 (0.9)
0.06					95.7 (0.9)
0.07					94.7 (0.8)
0.08					94.7 (0.8)
0.09					95.1 (0.8)
0.10					92.6 (0.8)
0.11					92.8 (0.8)

Table 14 continued

	W	W_c	W_{mp}	W_*	W_a
0.12					92.8(0.8)
0.13					92.8(0.8)
0.14					89.5(0.8)
0.15					89.5(0.8)
0.16					89.5(0.8)
0.17					85.0(0.8)
0.18					85.0(0.8)
0.19					85.0(0.7)
0.20					79.4(0.7)
δ	$q = 100, m = 39$				
0.01	93.9(0.9)	94.6(0.9)	94.6(0.9)	94.6(0.9)	94.6(0.9)
0.02					95.7(0.9)
0.03					95.7(0.9)
0.04					94.7(0.9)
0.05					94.7(0.8)
0.06					95.1(0.8)
0.07					92.6(0.8)
0.08					92.8(0.8)
0.09					92.8(0.8)
0.10					89.5(0.8)
0.11					89.5(0.8)
0.12					85.0(0.8)
0.13					85.0(0.8)
0.14					85.0(0.8)
0.15					79.4(0.8)
0.16					79.4(0.7)
0.17					79.4(0.7)
0.18					72.5(0.7)
0.19					72.5(0.7)
0.20					64.7(0.7)

Table 15 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the coverage probability and average length (in parentheses) of confidence intervals with nominal level 95% derived from profile, conditional, modified profile, penalized (Firth) and penalised based on adjusted responses log-likelihood ratios, with all coverage probabilities multiplied by 100

	W	W_c	W_{mp}	W_*	W_a
δ	$q = 1000, m = 1$				
0.31	0.0(0.4)	95.0(0.3)	6.3(0.3)	6.3(0.3)	34.4(0.3)
0.32					39.8(0.3)
0.33					48.2(0.3)
0.34					56.7(0.3)
0.35					62.2(0.3)
0.36					70.0(0.3)
0.37					77.0(0.3)
0.38					81.1(0.3)
0.39					86.3(0.3)
0.40					89.1(0.3)
0.41					92.3(0.3)
0.42					94.6(0.3)
0.43					95.5(0.3)
0.44					96.1(0.3)
0.45					96.0(0.3)
0.46					95.7(0.3)
0.47					94.8(0.3)
0.48					92.5(0.3)
0.49					90.5(0.3)
0.50					88.0(0.3)
δ	$q = 1000, m = 3$				
0.06	4.1(0.3)	95.0(0.3)	91.2(0.3)	91.2(0.3)	34.4(0.3)
0.07					42.5(0.3)
0.08					53.9(0.3)
0.09					62.2(0.3)
0.10					70.0(0.3)
0.11					77.0(0.3)
0.12					83.0(0.3)
0.13					87.7(0.3)
0.14					91.3(0.3)
0.15					94.5(0.3)
0.16					95.8(0.3)
0.17					96.1(0.3)
0.18					95.5(0.3)
0.19					94.0(0.3)
0.20					92.5(0.3)
0.21					89.3(0.3)

Table 15 continued

	W	W_c	W_{mp}	W_*	W_a
0.22					85.0(0.3)
0.23					81.5(0.3)
0.24					75.3(0.3)
0.25					68.1(0.3)

Table 16 Binomial matched observations with true log odds ratio $\psi = 1$. Second to sixth columns show the coverage probability and average length (in parentheses) of confidence intervals with nominal level 95% derived from profile, conditional, modified profile, penalized (Firth) and penalised based on adjusted responses log-likelihood ratios, with all coverage probabilities multiplied by 100

	W	W_c	W_{mp}	W_*	W_a
δ	$q = 1000, m = 11$				
0.01	79.1(0.3)	95.0(0.3)	95.0(0.3)	95.0(0.3)	86.2(0.3)
0.02					91.2(0.3)
0.03					94.1(0.3)
0.04					95.4(0.3)
0.05					94.9(0.3)
0.06					92.4(0.3)
0.07					87.9(0.3)
0.08					81.5(0.3)
0.09					75.3(0.3)
0.10					65.6(0.3)
0.11					54.6(0.3)
0.12					43.3(0.3)
0.13					35.1(0.2)
0.14					25.1(0.2)
0.15					16.9(0.2)
0.16					12.0(0.2)
0.17					7.2(0.2)
0.18					4.0(0.2)
0.19					2.4(0.2)
0.20					1.2(0.2)
δ	$q = 1000, m = 39$				
0.01	93.6(0.3)	95.0(0.3)	95.0(0.3)	95.0(0.3)	95.0(0.3)
0.02					94.1(0.3)
0.03					91.3(0.3)
0.04					86.5(0.3)

Table 16 continued

	W	W_c	W_{mp}	W_*	W_a
0.05					79.5 (0.3)
0.06					70.6 (0.3)
0.07					57.4 (0.3)
0.08					46.1 (0.3)
0.09					35.1 (0.3)
0.10					25.1 (0.2)
0.11					16.9 (0.2)
0.12					10.6 (0.2)
0.13					6.2 (0.2)
0.14					2.9 (0.2)
0.15					1.4 (0.2)
0.16					0.6 (0.2)
0.17					0.3 (0.2)
0.18					0.1 (0.2)
0.19					0.0 (0.2)
0.20					0.0 (0.2)

Table 17 Binomial matched observations with true log odds ratio $\psi = 1$. The values of δ that minimize the bias of the penalised log-likelihood estimator based on adjusted responses for various values of q and m . δ was chosen from a set of 50 values ranging from 0.01 to 0.50

m	$q = 4$	$q = 8$	$q = 12$	$q = 16$	$q = 20$	$q = 24$	$q = 28$	$q = 32$
1	0.05	0.37	0.43	0.45	0.45	0.45	0.45	0.45
3	0.01	0.12	0.17	0.18	0.18	0.18	0.18	0.18
5	0.01	0.06	0.10	0.11	0.11	0.11	0.11	0.11
7	0.01	0.04	0.08	0.08	0.08	0.08	0.08	0.08
9	0.01	0.03	0.06	0.07	0.07	0.07	0.06	0.06
11	0.01	0.02	0.05	0.06	0.06	0.06	0.05	0.05
13	0.01	0.02	0.05	0.05	0.05	0.05	0.05	0.05
15	0.01	0.01	0.05	0.05	0.05	0.05	0.04	0.04

Acknowledgements The author is deeply indebted to Professor Ioannis Kosmidis for his many insightful remarks and critical comments and suggestions which have led to significantly improving the quality and presentation of the results.

Funding This work was supported by the University College London EPSRC funded studentship in Mathematical Sciences.

Declarations

Supplementary information Not applicable

Conflict of interest The author declares that she has no Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barndorff-Nielsen OE (1983) On a formula for the distribution of the maximum likelihood estimator. *Biometrika* 70(2):343–365. <https://doi.org/10.1093/biomet/70.2.343>
- Breslow N (1981) Odds ratio estimators when the data are sparse. *Biometrika* 68(1):73–84. <https://doi.org/10.1093/biomet/68.1.73>
- Butler K, Stephens MA (2017) The distribution of a sum of independent binomial random variables. *Methodol. Comp. Appl. Prob.* 19:557–571. <https://doi.org/10.1007/s11009-016-9533-4>
- Cox DR, Reid N (1987) Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society* 49(1):1–39. <https://doi.org/10.1111/j.2517-6161.1987.tb01422.x>
- Cox DR, Snell EJ (1989) *Analysis of Binary Data*. Chapman and Hall, London
- Davison AC (1988) Approximate conditional inference in generalized linear models. *Journal of the Royal Statistical Society* 50(3):445–461. <https://doi.org/10.1111/j.2517-6161.1988.tb01740.x>
- Davison AC (2003) *Statistical Models*. Cambridge University Press, Cambridge
- Firth D (1993) Bias reduction of maximum likelihood estimates. *Biometrika* 80(1):27–38. <https://doi.org/10.1093/biomet/80.1.27>
- Florescu I (2014) *Probability and Stochastic Processes*. John Wiley and Sons, New Jersey
- Gart JJ (1970) Point and interval estimation of the common odds ratio in the combination of 2×2 tables with fixed marginals. *Biometrika* 57(3):471–475. <https://doi.org/10.2307/2334765>
- Gart JJ (1971) The comparison of proportions: A review of significance tests, confidence intervals and adjustments for stratification. *Int. Statist. Rev.* 39(2):148–169. <https://doi.org/10.2307/1402171>
- Gourieroux C, Monfort A, Renault E (1993) Indirect inference. *Journal of Applied Econometrics* 8:85–118. <https://doi.org/10.1002/jae.3950080507>
- Kosmidis I, Firth D (2021) Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models. *Biometrika* 108(1):71–82. <https://doi.org/10.1093/biomet/asaa052>
- Kuk AYC (1995) Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society* 57(2):395–407. <https://doi.org/10.1111/j.2517-6161.1995.tb02035.x>
- Lunardon N (2018) On bias reduction and incidental parameters. *Biometrika* 105(1):233–238. <https://doi.org/10.1093/biomet/asx079>
- Magnus JR, Neudecker H (2019) *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd edn. John Wiley and Sons, New Jersey

- McCullagh P, Tibshirani R (1990) A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society* 52(2):325–344. <https://doi.org/10.1111/j.2517-6161.1990.tb01790.x>
- Neyman J, Scott EL (1948) Consistent estimates based on partially consistent observations. *Econometrika* 16(1):1–32. <https://doi.org/10.2307/1914288>
- Pace L, Salvan A (1997) *Principles of Statistical Inference: from a Neo-Fisherian Perspective*. World Scientific Publishing Co, Singapore
- Sartori N (2003) Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika* 90(3):533–549. <https://doi.org/10.1093/biomet/90.3.533>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.